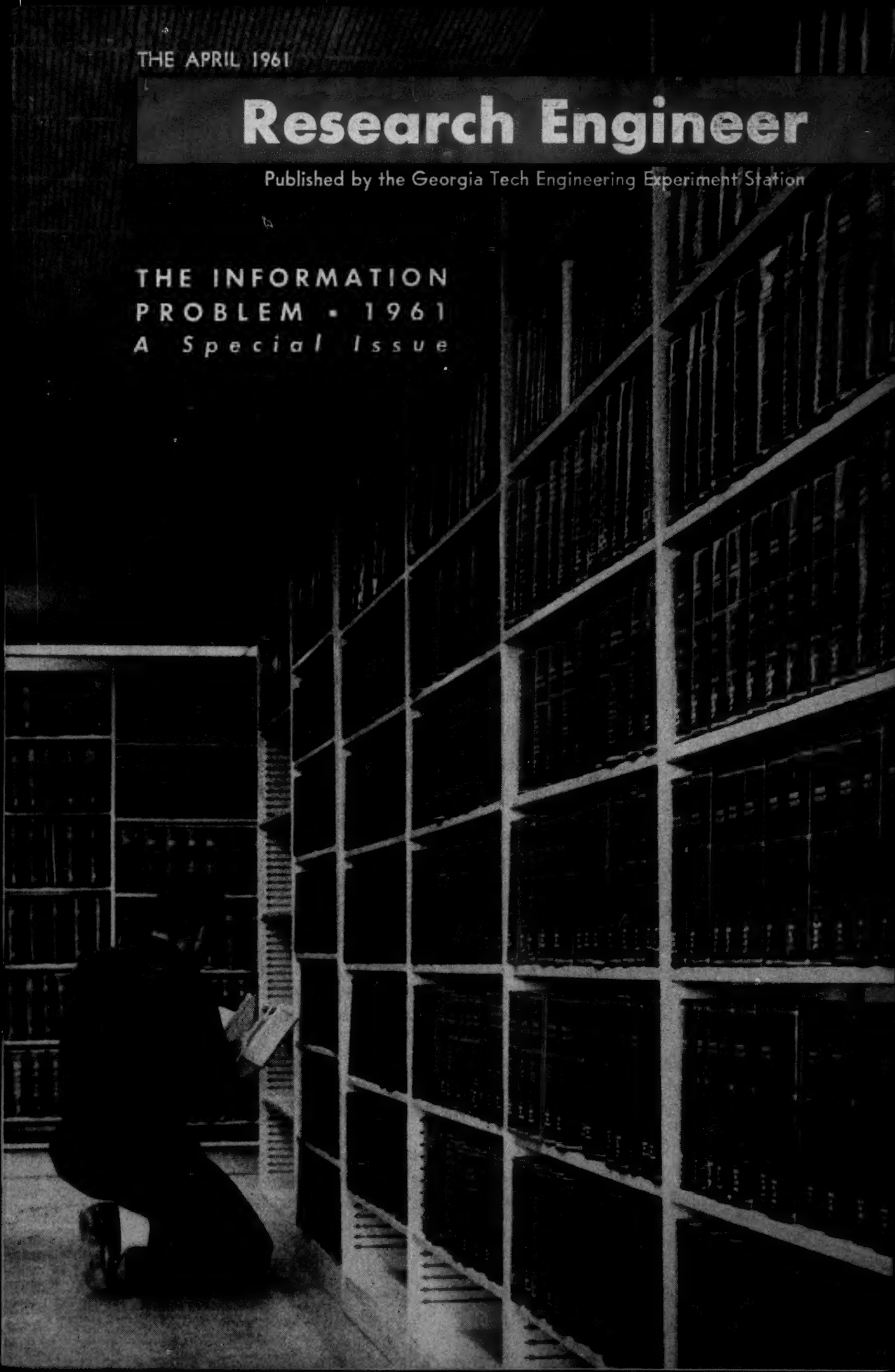
THE APRIL 1961

Research Engineer

Published by the Georgia Tech Engineering Experiment Station

THE INFORMATION PROBLEM • 1961
A Special Issue



RESEARCH ENGINEER

VOLUME 16, NO. 2 APRIL, 1961

Published five times a year by the Engineering Experiment Station
Georgia Institute of Technology, Atlanta, Georgia

contents

the cover

The expanse of technical information stored in shelves and shelves of books, seem to almost transcend man.

Cover photo by Bill Diehl Jr.

THE RESEARCH ENGINEER is published five times a year in February, April, June, October and December by the Engineering Experiment Station, Georgia Institute of Technology. Second-class postage paid at Atlanta, Georgia.

ONE DRIVING FORCE that allowed the United States to overtake the countries of Europe is similar to the one that enabled Russia to surpass this country in many areas during the past decade. In the foreseeable future, other countries may rise up and equal or surpass the technology and consequently the standard of living of both the United States and Russia. This driving force that permits whole countries to progress in leaps and bounds over a relatively short period of time is simply utilization of *accumulated technical information.*

Its importance cannot be overestimated. Its results include not only progress, but savings in time, costs, and effort.

How many inventors have spent years in the discovery of new knowledge only to find that the knowledge was several years old or published in another language?

How many scientists and engineers have put forth effort and expended funds toward the development of an application and then find duplicity of effort?

Availability of technical information also assists us in another way. The time lag between discovery of basic knowledge and its application can be greatly reduced. Thus, reduction of time lag, savings of time, costs, and effort and progress are greatly influenced by the ability of a country to retrieve and disseminate valuable technical information.

As the articles in this issue indicate, Georgia Tech is doing its share of this necessary activity. Its library, department records, and research facilities all contribute continuously in this area to the State, region, and the country.

E. D. Harrison *President*

TECHNICAL INFORMATION . . . scientist's utility

by Robert J. Kyle, Head Technical Information Section

DESPITE the flood of papers written about the information problem in recent years, few people realize that technical information is now big business. Yet, each year roughly $300,000,000 is spent in the United States alone just supplying the information needs of research and development.

To the scientist and engineer, technical information is a utility like electricity or water is to the home owner. And, like municipal utilities, some information services are vastly superior to others. But, almost all are worth far more than their cost.

New companies to provide information services are being organized at a surprising rate. The situation has progressed to the point that speculative investors are seeking companies which hold promise of developing a strong position in this field called technical documentation.

Technical documentation ranges from the recording of scientific facts through the storing, relocating, and the distribution of the information as an answer to a question. It includes the dissemination of information to keep people abreast of recent developments and the retrieval of information to meet specific needs.

Recent developments in computer processing, photographic recording, and other techniques have glamorized the information field. The technical documentalist is provided with exotic tools interesting to scientists and laymen alike.

The most impressive equipment combines computers with very small pictures of the original documents. The photographic copies (or *storage*) permit the filing of collections of information in very small volumes. Specialized libraries can be recorded in boxes which only line part of a wall in a medium-sized room. Some of these machines can select references which contain desired data and produce a full size copy.

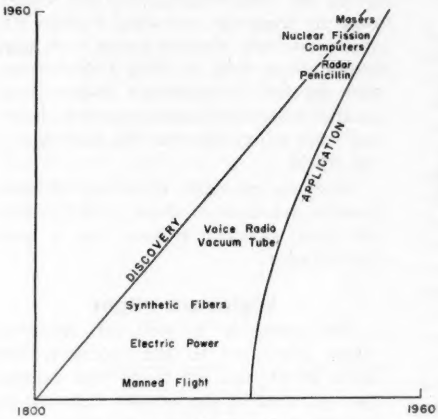
Difficulties still exist. The biggest bottle-necks in information processing are found in indexing, abstracting, and translating. It's hard to think of a machine studying an article and preparing an abstract. Yet, fairly reasonable abstracts are being prepared by machine already, and some progress has been made with machine indexing. Translation by machine is receiving a great deal of attention, and useful results will be economically feasible within a few years.

The task of supplying information to computers is another problem. Work, costs, and errors in retyping journals and books on magnetic tape or other computer input are overwhelming. But before long machines to read printed pages will alleviate this problem. Some machines which can read a single variety of type are available, and more versatile machines are to be unveiled this year.

Thus, the future looks bright. But the equipment discussed is very expensive. It can only be used in cases where many thousands of inquiries are to be answered each year. Even the most sophisticated techniques can serve only limited subject areas. The day when a library like Georgia Tech's can be automated is still not in sight.

There are many information techniques suitable for small-scale use. They, too, are worthy subjects for discussion. These systems and the traditional techniques play the major role in information processing today. They must be extended and so must the glamorous documentation services if we are to keep the information flowing.

To fail to satiate the country's technical information needs will surely slow its technological advances. For the country's own good, and its position in the present world-wide scientific competition, support must be given to research and application in the field of documentation.



THE SQUEEZE
Figure 1

THE INFORMATION EXPLOSION

by

Lawrence W. Ross

EVERY 24 HOURS enough technical papers are produced around the world to fill seven sets of the Encyclopedia Britannica. One man, scanning the 60 million pages of technical literature produced last year, would need 465 years to finish the job.

The worst is yet to come. The total quantity of technical literature in existence doubles every ten years. Scientific expenditures are going up even faster, increasing at 10 per cent annually, which will eventually bring even more emphasis to scientific publication.

Scientist's Dilemma

What can the individual do in the face of this flood of facts? He is under obligation to keep pace with advances. He needs to be able to retrieve all facts about a particular subject whenever he contemplates new work.

He cannot meet these demands any longer, unless he is engaged in a particularly narrow specialty. Already several legendary tales have made the rounds, telling of six-figure sums spent on work already completed and reported by others.

Figure 1 reveals still another pressure upon the poor scientist: the necessity to translate fundamental discoveries into practical products with ever-decreasing delay.

New Outlooks

When scientists cannot keep pace, new ideas and outlooks are called for. Our traditional means of scientific communication are plainly outmoded. Once upon a time ideas were *communicated* instead of *disseminated* like pollen on the breeze. That was when a well-equipped scientific library consisted of a single bookrack (as the Royal Society library did 120 years ago), and when a scientist could count on understanding the work reported in other branches of science.

At the turn of the century, abstract journals such as *Chemical Abstracts* appeared to serve the needs of scientists

who could not keep tract of the growing number of journals. Now there are about 300 such abstract journals serving all major fields of science.

There the matter rests. We have not improved upon it in 60 years. Science has come under increasing criticism for, in Jacques Barzun's words, "insisting pedantically on one restricted means of imparting new-found truth."

It has been suggested that scientific articles are out of date, and that abstracts should become the standard medium of scientific communication. Papers would still be produced as they are now, according to these schemes, but they would not be published.

Instead, the abstract would be published. The man who really wants to read the article then is able to order a copy, and the libraries are spared the necessity of storing it and the thousands of other papers which may be read only seldom, or only once or twice, or never.

As we shall see, this system of information-by-abstract is already in use among many information-consuming groups.

Modus Vivendi

The Latin phrase *modus vivendi* describes the relationship now existing between scientists and their literature. They are getting along the best they can.

Most scientists regularly read a handful of pet journals. The physicist, for example, typically reads *Physical Review* (or scans it, more likely), *Journal of Applied Physics, Review of Scientific Instruments* and a couple more relating to his specialty (e.g., *Acta Crystallographica*).

He has two more regular sources of information: *Reviews of Modern Physics* and the *Bulletin* of the American Physical Society. The first gives him the opportunity to learn what is going on in other fields of physics. The *Bulletin,* whose importance cannot be overemphasized, reports the papers to be presented at the Society's conventions, and thus fills a vital need: keeping physicists informed of the activities of their counterparts around the nation.

A conscientious physicist may attempt to fill the remaining gaps in his knowledge by regularly reviewing *Physics Abstracts* or other abstract journals. A large organization such as Bell Laboratories, with its *Bell Laboratories Index,* serves its staff with still another means of learning what is published in the literature of the world.

Now our physicist, if he uses all these sources, can learn of about 20,000 articles per year, although it costs him a great deal of time.

Engineer's Plight

Our physicist is well off, however, when compared to the engineer. The areas of physics are fairly well defined; two per cent of the world's technical literature may cover about everything of interest to physicists working in highly theoretical areas.

Not so in engineering science! Chemical Engineering, for example, which is the vocation of your author, overlaps a very large number of scientific disciplines. Practically any technical journal in the world could contain material of interest to some chemical engineer somewhere.

Engineers have still another handicap, in your author's experience. Their employers are often corporations whose executives regard reading time as nonproductive time. This brings unmistakable pressure to minimize "book time."

The Future of Information

The much-heralded *population explosion* will be a small blast compared to the information explosion. After all, population is increasing only two per cent annually. What happens if this rate continues indefinitely? Of course, it is absurd to contemplate results obtained on this basis.

In the same way, information cannot always increase exponentially, or the

results are equally absurd: after 200 years at this rate, the number of scientists and publications would both increase a millionfold.

There is a much better way out of the riddle, suggested by Russian investigators. If we assume that the growth of printed literature is indeed exponential, we have the sort of curve shown at left in Figure 2. But if we assume that science will eventually become saturated with literature, there will be a levelling-off of the sort shown by the right-hand curve (logistic curve). Here is a hypothesis we can test.

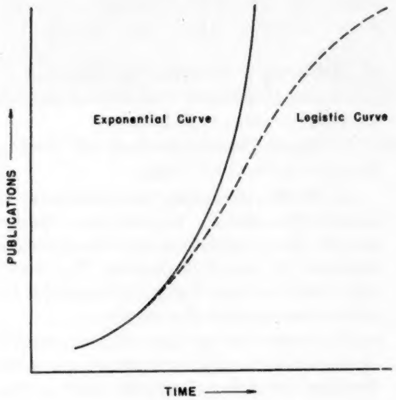


Figure 2

In limited areas of science a levelling-off has taken place. This gives weight to the hypothesis, but does not yet allow us to draw conclusions concerning all of science.

Of course, a cataclysm such as a world war would alter all assumptions. World War II, however, bad as it was, caused only a wiggle in the upward progress of most publications. In chemistry there was a decline in publications, but in physics the rate was steadily up (although reduced), and in journals devoted to scientific instruments there was an uninterrupted exponential increase.

Predictions

The information problem will get much worse before it gets better. The really large population centers (China, India) have only begun their scientific output, South America is just now ready to enter the competition, and Africa will develop last of all. Until all the world is producing its share of scientific literature, the increase will tend to be exponential with time.

In this writer's opinion, scientists will soon have to rely on others to collect their facts for them. This plainly means that information services will become more and more prominent.

But new methods of disseminating information are needed. The present 18-month lag between submission of a paper and its publication is already unacceptable, as Figure 1 emphasizes. Everyone is dissatisfied with the journals, but tradition is strong.

There is one way in which all three elements—speed, sheer volume, and tradition—could be handled satisfactorily. Suppose that all papers to be published under society auspices are forwarded to a central agency (for example, a federation of scientific societies) and *announced* by a publication of that agency. An abstract might well accompany the announcement of title.

The author would receive publication credit, thereby satisfying both tradition and his superiors. Editing time could be reduced to the point where the paper is officially accepted, thereby introducing speed and economy. Volume would be eliminated, at least to the point where one page would account for several articles. A speedy system for supplying reprints at a cost which would support the plan would satisfy the scientists who need the information.

Unless some such plan is adopted, we face *information inflation*. When there is too much currency in circulation, it is debased, and there is inflation. Information, the currency of science, can become inflated in the same way. The end is the same: collapse of the system.

And then the Search began

by Francis Kaiser

EVEN IF a man had the time to read the best of the world's literature in his fields, he would be stymied because less than half of it is in English, more than one-fourth in French or German, and more than one-third in Russian, Japanese, or Chinese. Most American scientists are ill-prepared to read even the German or French.

To meet the need for foreign material, a shotgun approach to translation has been used in recent years. Free lance translators, commercial translating firms, industrial and government translating sections, publishers subsidized by government or professional societies issuing cover-to-cover translations of journals, and machine translation research have evolved from this approach.

Making a Translation Search

The most efficient way to search for a translation today is as follows:

1. Search published lists of existing translations in the Library.

2. Write, telegraph, or telephone to translation centers, to find out whether or not the translation is very recently-received or is still in process. The Georgia Tech's Price Gilbert Memorial Library also supplies this service.

3. Secure the services of a translator, if the above search is fruitless, using the facilities of an organization such as the Technical Information Section, Engineering Experiment Station.

At Georgia Tech, the first two steps are usually handled by the Library. As one of seven depositories for printed copies of technical translations issued by the Office of Technical Services, U. S. Department of Commerce, the Library has received more than 6,000 translations during the past two years. In addition, it subscribes to more than forty cover-to-cover translations of foreign journals and receives hundreds of translations annually from other sources, such as the Atomic Energy Commission and foreign universities and governments. The Library may draw also upon the holdings of other libraries and research organizations if necessary.

Photo Credit: Harley Ferguson

TECHNICAL TRANSLATION IS NECESSARY FOR SCIENTIFIC DEVELOPMENT.

Translation Centers in the U. S.

An international network of translation centers is developing in the United States and Europe at present. In America, translations from all sources are channelled into the following two centers, which exchange all items with one another and cooperate in abstracting, indexing, and distributing them to the public: (1) the Special Libraries Association Translations Center, housed in the John Crerar Library, Chicago, Illinois, which collects translations from nongovernmental sources, domestic and foreign, and (2) the Office of Technical Services, U. S. Department of Commerce, Washington, D. C., which collects them from government sources, domestic and foreign. They jointly compile and publish *Technical Translations,* a bi-weekly abstract journal which gives complete information about their translations.

Foreign Translation Program

In 1958, the U. S. Government embarked on a foreign translation program, and appropriated $1,200,000 for translation work in certain countries. To date, contracts under this program have been signed with Israel, Poland, and Yugoslavia, calling for the translation of over 50,000 pages of text from Russian and other East European languages. The completed translations will be edited, printed, and sold in the United States through the office of Technical Services, U. S. Department of Commerce.

International Translation Centers

As further impetus for rapid exchange of translations between Europe and the United States, two international translation centers have been established in Europe within the past six months. The first, known as the European Translation Centre, was created in October, 1960, by the twelve member countries of the European Productivity Agency (EPA). Its purpose will be to notify industry and research of all existing Western European language translations from Russian and East European scientific and technical literature, and to make available those which cannot be obtained through commercial channels. The Centre will be housed at the Library of the Technical

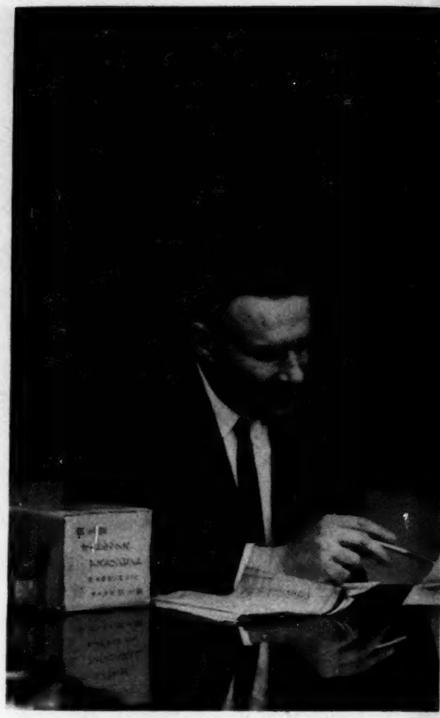
Institute of Delft, Netherlands. Its holdings will be listed in *Technical Translations*.

Not to be confused with the European Translation Centre is another new center, Transatom, established in Brussels, Belgium, in December, 1960, by a joint agreement of the European Atomic Energy Community (EURATOM), the United Kingdom Atomic Energy Authority, and the U. S. Atomic Energy Commission. Transatom will be primarily concerned with creating and maintaining a master index file of translations of nuclear literature, from which it will publish a monthly *Transatom Bulletin*. Information will also be exchanged with the European Translation Centre, at Delft, Netherlands, to prevent duplication of activities.

Made-to-Order and Commercial

If the existing translation is not locatable, the scientist and translator work together informally, as a team. For this reason, many companies and government agencies have a staff translator or a translating section attached to the research department. At Georgia Tech, a similar personalized translating service is provided by the Technical Information Section of the Engineering Experiment Station which maintains a roster of translator talent available on or near the campus. The translator usually meets with the scientist and they go over the material together. Thus, it is possible to skip nonpertinent sections or even eliminate whole articles which prove to contain no useful information. Another advantage of this method is that the scientists are able to ask questions if the translator makes a wrong choice of synonym for a key word. These "face-to-face" translations are much cheaper than custom-made written translations. The average cost is $2.25 per hour for the translator's time, plus any incidental expenses incurred in securing a copy of the original article, added to the cost of the requestor's own time.

When no campus translator can be found for a given job, the work is often



THIS SCIENTIST IS ASSISTED BY USE

assigned to a commercial translating firm. There are no professional standards or licensing requirements for translators in the United States. Therefore, the prospective customer must exercise care to secure references and to investigate the skill of a translator before placing an order. The Special Libraries Association is modifying the directory of translators first published in 1959. *Translators and Translations: Service and Sources*, edited by this paper's author, lists a few of the many firms and individuals available, without guaranteeing the claims made by those listed.

Commercial translators offer two types of service: (1) made-to-order, or "cus-

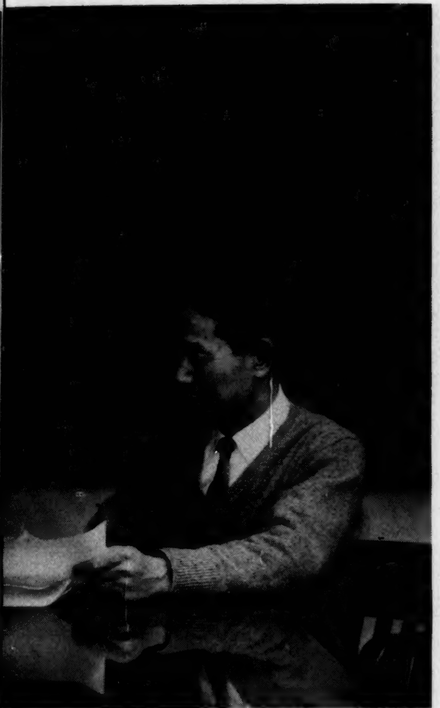


Photo Credit: Harley Ferguson

OF "FACE TO FACE" TRANSLATION.

tom-made" translations, which become the exclusive property of the purchaser, and (2) printed translations, which remain the property of the translator, copies being sold to as many purchasers as request them. The cost of made-to-order translations is relatively high, as the following rates of typical firms will show:

Rates Charged for Custom-Made Translations — Rates Per Thousand Words English

Language Translated	Firm A	Firm B	Firm C	Firm D
German	*$18*	*$15*	*$8*	*$15*
Russian	*$19*	*$18*	*$8*	*$20*
Japanese	*$20*	*$20*	*$20*	*$35*

The second type of translation if available, is much less expensive. The costs of typical printed, *i.e.*, mass circulation translations from commercial firms vary from about $1.50 to $4.50 per 1000 words of English. The price depends on the policies of the translator and the number of copies which will be sold.

Some firms are able to complete orders in five to seven days. Others, according to the rarity of the language and the length of the article, may take four to six weeks. In general, a firm which specializes in a given subject field or in a particular language can give faster service than one which handles many diverse subjects and a large number of languages.

Future Trends

The entire field of translation is in transition. We have progressed beyond the pioneer stage when, because of the few translations available, nearly every article had to be translated to order. However, we still have a long way to go before reaching an ideal solution. Experts differ, but all agree that we will continue to need both men and machines. Translators of the future will be a highly-trained, select group.

For the present, we must make the most of available facilities. The National Science Foundation, which is investing over a million dollars per year to subsidize translations and translation research, is currently supporting a nationwide survey of translation activities in the fields of science and technology. The results of the survey, to be completed in mid-1961, should lead to more effective use of the translation program in the future.

Within the next few years the transition from human to machine translation will begin. Ultimately, machine translations will reduce the cost and time of translations so markedly that language barriers to information exchange may become almost negligible. Already, the military incentive, which gave the original impetus to large-scale translation programs, is being replaced by general scientific, industrial, social, and even political applications.

Tech's Price Gilbert Library in today's technological society plays

A NEW ROLE

by Graham Roberts

Photo Credit: L. C. Prowse

THE NEW SCIENTIFIC ERA ushered in by World War II has brought about a revolution in the concept of the role of the technical library. The volume of research, both pure and applied, and the resulting publication have increased at an astronomical rate.

The insatiable demands of today's research have destroyed the handbook-textbook concept in technical libraries. The present-day technical library, whether it be academic or industrial in nature, must be research-oriented and must be organized to locate needed information in the most selective and efficient manner. On the campus, such libraries have developed into storehouses of information geared not only to the factual needs of the undergraduate but also to the research needs of the graduate student, the teaching and research faculties, and their industrial clientele. The Georgia Tech Library is an example of the revolution which has taken place. The growth in its physical plant and facilities, the steady development in the services and caliber of its staff, the rapid increase of its resources of books, magazines and reports and their accompanying indexes, abstracts and bibliographies—all these provide concrete evidence of the new role played by the Georgia Tech Price Gilbert Memorial Library. Completed in 1953, this modern building is an air-conditioned contemporary structure of five floors designed to hold approximately 500,000 volumes and 800 people. Use of glass and color, a warm and friendly atmosphere, the general feeling of spaciousness and freedom of movement, and use of the best in furniture and equipment symbolize the new technological environment.

Growth and specialization in teaching and research programs have led to a corresponding growth and specialization in library services as the library has tried to keep pace with the institution's accelerated demands for the acquiring, processing and servicing of materials and information. To meet the challenge a gradual expansion of the library staff has been

Georgia Tech's Library—Cont.

necessary. From a total of 10 persons in 1946, the staff has increased to 41 by 1961. Special service areas for General Studies and for Science Technology have been established and are manned by professional librarians 84 hours per week. The library is open for study a total of 90 hours per week. Informational services on such special collections as patents, maps and technical reports are handled by staff members particularly competent in these areas.

During the past fifteen years, the number of volumes has grown from 77,000 to 240,000. However, it is the quality of the expansion far more than its physical volume which is of primary importance. In this short period, the library's collection has developed into one of both regional and national importance. Last year a total of 4,621 interlibrary loan requests were received from borrowers in 41 states and 8 foreign nations. A further illustration of growth and development is the fact that ten years ago the library did not subscribe to a single Russian journal. Today it subscribes to 82 scientific and technical journals either in Russian or translated from Russian.

Researchers from all over the nation seek our research information at Georgia Tech. The current serial subscriptions, including newspapers, periodicals, transactions and yearbooks now number about 7000. Of these over 2800 are scientific and technical journals, half of which come from foreign lands. Reference books, monographs and textbooks are available in plentiful supply, but it is the current periodical which constitutes the core of the present day technical library.

Several special collections also add to Tech's library resources. One is the collection of *technical reports*. The technical report is often the first place that the results of applied research are published. These reports which vary in length from a few pages to several volumes are the important product of the ever-expanding government—industry contract system which began to develop during World War II. The library now has 75,000 reports. It is a depository for publications issued by such agencies as the Atomic Energy Commission, the Rand Corporation, the Army Map Service, the Office of Technical Services of the U. S. Department of Commerce as well as many outstanding American and foreign research laboratories, and experiment stations connected with educational institutions.

The library's collection of patents is another significant source for the researcher. Since May 7, 1946 the library has received copies of all U. S. patents. These include the complete specifications and accompanying drawings of all patents beginning with patent number 2,399,611. The recent acquisition of an index to the classification system used by the U. S. Patent Office now enables the Library to identify all patents by class and sub-class which were issued between 1790 and September 1959. Also available is a complete file of the OFFICIAL GAZETTE OF THE UNITED STATES PATENT OFFICE and a long run of its British equivalent, the ABRIDGED BRITISH PATENT SPECIFICATIONS. Current subscriptions are held to patent journals of Canada, England, Germany, Australia and Russia. The Tech Library has become a southeastern regional center for patent information and is providing service for an increasing number of researchers, patent lawyers, and aspiring inventors.

As library collections have become larger and more complex and as research needs have placed a premium on the quick and accurate retrieval of information, those tools (abstracts, indexes, bibliographies, etc.) which serve as keys to the information have become all important in the technical library. The Tech Library has responded with its policy of acquiring all important bibliographic tools needed to further the institution's teaching and research programs. The end result is that today, Tech's

Photo Credit: Harley Ferguson

A SCIENTIST RECEIVES HELP WHEN CONDUCTING HIS RESEARCH SURVEY.

collection of scientific and technical bibliographies is without peer in this region of the country. Without these keys to the books, periodicals and reports, the retrieval of information would be difficult.

Tech's Library also participates along with the other major university and research libraries in the Farmington Plan for the acquisition of foreign publications. Under this arrangement the Tech Library receives copies of all items in the field of textile engineering on a world wide basis. This program has added materially to the strength of the research resources available to the nation's textile engineers. Similar strong collections are being developed in those fields in which the institution offers the doctorate: Aeronautical Engineering, Chemical Engineering, Chemistry, Electrical Engineering, Physics, Civil Engineering, Sanitary Engineering, Mechanical Engineering, and Industrial Engineering.

In the past two decades, the Georgia Tech Library has expanded into one of the major technical and scientific collections of the nation. Georgia Tech's education and research is ably assisted by its library building and the collection which it houses.

A peaceful world, a progressive world, or certainly a more knowledgeable world will result from new

MACHINE TRANSLATION

by Robert W. Hays

THROUGHOUT HISTORY, man has solved his Tower of Babel problem—language translation—in only one way: he has trained translators. Recently, a new approach using electronic data processing equipment has been showing promise of accelerating translation.

No one knows for certain how many languages man uses in communication. The differences between accents, dialects, and languages are often difficult to specify. Yet, man must currently be using well over two thousand languages. Fluency in any two may require months or even years for a human translator, and at best the translation process is tedious and time-consuming. However, machine translation may permit exchanging ideas at high speed in many pairs of languages. *Machine translation*, abbreviated as MT, would meet many needs. Those research personnel who must now translate materials could be partly relieved of tedious work. The high cost of the technically trained translator could be lessened. Thus, much more scientific and technical literature would be available in translation.

MT may also aid in high-speed abstracting and information retrieval. A computer which translates one language into another will change a *natural language* (one written by people) into electrical charges, magnetic fields, blips

ONE OF THE "MT" SEMINARS FOR GEORGIA TECH RESEARCH MEMBERS.

on a photograph, or holes in a card. Such a language transformation may help simultaneously in abstracting or indexing an article for purposes of information retrieval.

The use of high speed electronic equipment has almost become commonplace. A monthly statement from a department store or utility company may have emerged from a data-processing center, and the procedure would have resembled MT. A human being programs a computer. The computer matches the customer's name against a file (almost like looking in a dictionary). It calculates his purchases and discounts, etc. (somewhat like deciding which words are subject, verb, etc.). It organizes the data (similar to syntactical arranging the target language sentence). And it prints a statement (analogous to typing a translation).

Because other data-processing resembles translation, translation might not seem difficult. A computer can connect a German word with an English word, as in relating *Wasser* and *water*. An example of word-for-word translation appears below in the first two lines of *Die Lorelei* by Heine

(I) (white, clean) (not) (what, which, something) (?) (it, so, they) (to inform, to set right, to signify, (that) (I)

(so, thus, in such a manner, like this) (mournful, sad, dismal) (?);

The German version is

Ich weiss nicht was soll es bedeuten,
Dass ich so traurig bin;

A free translation of Heine's lines would be:

I do not know what it means
That I am so sad;

The free translation deviates far from the word-for-word translation. One word had a *wrong* listing, and two words had no listings in the small dictionary used. A larger dictionary would perhaps show that *weiss* is not only an adjective (*white* or *clean*) but also a present-tense form of the verb *wissen* (*to know*); and that *bin* is a present-tense form of *sein* (*to be*). However, a dictionary (or in MT terms, a *glossary* or *memory*) large enough to include these possibilities and a translation for *soll* would leave even more choices in the parentheses.

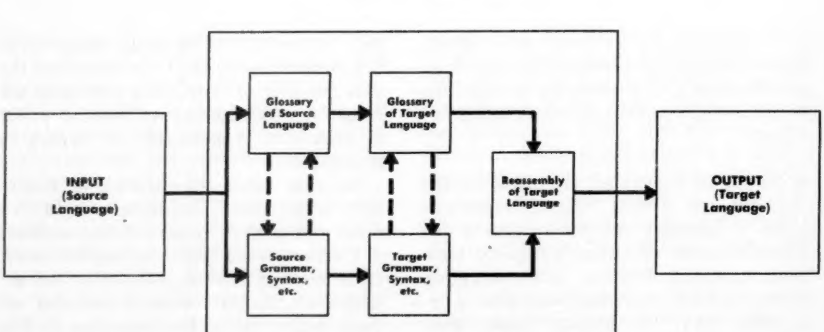
The fundamental barrier to word-for-word translation of MT is multiple meaning which forces the reader to choose from several possible meanings. (One desk-top dictionary gives over a hundred meanings for the English word *run.*) Even if MT supplied all conceivable meanings of each *source* (*input* or *foreign*) word, a human editor (*post-editor*) would have to choose among the meanings in the *target* (*output*) language. Although the choice seems fairly simple, it requires almost as much skill and time as does a conventional translation. Word-for-word translation is not sufficient.

To supplement the word-for-word translation of an *automatic dictionary,* more sophisticated programs are necessary. A simplified schematic of an adequate MT program might look like Figure 1.

The heart of this program is the grammar and syntax *operator,* for which

Figure 1

SCHEMATIC OF AN MT PROGRAM



NOTE: Broken Lines Indicate Interaction Feedback To Aid in Making Unresolved Decisions.

many different schemes have been proposed. The simplest would use a human pre-editor and post-editor to make choices too subtle for a computer. This plan seems largely to have been discarded. Other plans call for special *algorithms* (rules) for translating. Repeated *iterations* (passes) across the sentences to discover beginnings and endings of clauses and phrases, subjects and predicates of clauses, and other grammatical blocks will undoubtedly be necessary. These blocks will suggest the relations of the words which they contain.

Another plan calls for reducing the number of English equivalents for each source word. In translating *Die Lorelei, so, thus,* and *in such a manner* might reduce to *so* as a *cover word.* The output may however often not be *smoothed.* In other words the output may not be fluent, idiomatic English. Or, instead of a conventional dictionary a thesaurus, such as Roget's *Thesaurus,* with words grouped by meaning, might be better for the MT memory. When a word like *loss* occurred, the glossary would produce several thesaurus numbers. For the preceding word, *heat,* the glossary would produce several thesaurus numbers. The number common to the two words would suggest which meaning is intended.

An intermediate language with sets of vocabulary, morphological, and syntactical elements has been proposed, as has the use of artificial language (such as Interlingua or Esperanto). The most ambitious proposals call for a universal grammar embracing elements common to many languages, such as indicators of time, subjects of ideas, indicators of sex, etc. Thus, other languages, like English, can express the idea of *be.* Since the verb *to be* is suppressed in some languages, particular *operators* would be needed to supplement the general operator.

The many plans for bringing about MT result from the number of MT groups. At least eight countries have MT groups: The U. S., the U.S.S.R., England, Italy, Mexico, Communist China, Japan, and Israel. In the United States at least nine universities have groups interested in MT: Georgetown, Harvard, M.I.T., University of California, University of Michigan, University of Washington (two groups), Wayne State, University of Texas, and Georgia Tech. Corporate and governmental MT groups include the National Bureau of Standards, Ramo-Wooldridge, Rand Corporation, Arthur D. Little, Inc., Planning Research Corporation, International Business Machines Corporation, and Norair Division of Northrop.

Nearly all of the groups outside the U.S.S.R. and Communist China have received partial or entire U. S. governmental sponsorship, either from the National Science Foundation or from the U. S. armed services. English is the output language in nineteen programs and the source language in two programs. Russian is the source language for thirteen programs. Other languages include German, Polish, Italian, Chinese, Arabic, Japanese, French, Czech, and Serbo-Croatian.[1]

Russian progress in MT is hard to appraise. At least four major Russian groups are definitely working on MT, and the number of Soviet groups may be as high as sixty. The Russians need MT perhaps even more than does the United States, for the U.S.S.R. has many languages within its boundaries.

MT has evolved most rapidly in the last decade. In 1933, P. P. Troyansky, of the U.S.S.R., is supposed to have offered an MT scheme, but he never developed his project. Interest revived after World War II, but MT made few practical advances until 1948 and 1949 when Warren Weaver and A. D. Booth offered MT proposals. In 1950 and 1951 Erwin Reifler and Y. Bar-Hillel offered plans and status reviews for MT; and a machine translation conference met at M.I.T. in 1952.

In 1954 International Business Machines Corporation offered the first public demonstration of simple MT, and William N. Locke and Victor H. Yngve

of M.I.T. began publication of the periodical *Mechanical Translation*. Also in 1954, Anthony G. Oettinger received the Harvard PhD with a doctoral thesis on an automatic dictionary.

Since 1956, many MT groups have organized. A national symposium on MT met in 1960, and an international conference is scheduled in England this year.

Georgia Tech's MT group has functioned for about a year, with discussions of various problems in MT. Dr. B. J. Dasher, Director of the School of Electrical Engineering, has been compiling a structure-based grammar for English with elementary sentence patterns as components. The principles of this grammar may prove sufficiently general to apply to a large group of languages. Mr. James Gough, Jr., Assistant Professor of Modern Languages, has been making distribution studies of German *place-adverbs* and prepositional phrases of place. His ultimate aim is a computer sentence-generating routine.

Research groups have concentrated on the program for the machines, rather than on the computers themselves (the *hardware*). Most investigators have felt the equipment will be available by the time it is needed. The first step in the process—a machine to *read* type—seems almost ready. The last step—the *print* step—seems adequate now. The memories are probably large enough. Until recently, memory capacity has been a limitation.

The most serious problems are the operators for syntax, grammar, and reassembly. Languages have inherent difficulties. These difficulties include, for written English the following: homographs (such as *black* and *Black* or *lead* and *lead*), idioms (*to be free from*), word order, irregular plurals (*men, people*), inflectional endings (*dropPED* but *succeeDED*), prepositions (*on the table* but *against* or *on* the wall), grammatical and typographical errors, neologisms (new words), polysemy (multiple meaning), abbreviations and symbols, and perhaps even puns and slang (*real cool cat*).

Yet to a limited degree, MT is already a reality. At least Pidgin English can come out of the machine. Last year International Business Machines Corporation submitted to a congressional committee an MT treatment of Khrushchev's U-2 speech to the Supreme Soviet of the U.S.S.R. on May 7, 1960. One paragraph of the MT version read:

Authoritative expert commission, investigated shot down/put together aircraft, established on foundation analyzed documentary data that this American aircraft is special prepared air scout. This aircraft had problem intersect all territory Soviet Union from region Pamira to Stake peninsula for the purpose of intelligence military and industrial objects our country with the help of taking of photographs. Simultaneously on board aircraft, except/besides aircraft cameras, found other reconnaissance apparatus for uncovering radar network, determination place and frequency working radio station and other radio-technical means.[2]

What is the future of MT? MT of fiction or poetry—transforming a classic masterpiece in one language into a classic masterpiece in another—seems unlikely in the foreseeable future. A computer may never make translations to equal the Twenty-Third Psalm or Fitzgerald's *Rubaiyat*. But machine translations will improve, and consistently dependable and clear text seems in sight. The output will probably be neither smooth reading nor flawless grammar. But it will be unambiguous enough for the reader to understand the writer's facts.

BIBLIOGRAPHY

1. National Science Foundation, *Current Research and Development in Scientific Documentation*. Washington: Superintendent of Documents, April 1959, October 1959, and November, 1960.
2. International Business Machines Corporation, "Machine Translation of a Russian Newspaper," *IBM Research News*. Yorktown Heights, N. Y.: International Business Machines Corporation, May, 1960, p. 8.

Technical Information Services

by Lawrence W. Ross and Robert J. Kyle

IN ORDER to reduce the time which scientists must spend in the library, many organizations (and practically all research groups) have set up library research organizations. Sometimes a single librarian with no technical education fulfills this function. Other times it is an extensive organization with a number of specialists in the most important fields.

The Technical Information Section (TIS) is Georgia Tech's library research organization. Like other similar groups around the country, TIS keeps Tech's scientists posted on new information, collects old information, gives advice on indexing, arranges translations, and finds answers to technical questions.

Outside agencies also call on TIS. The range of activity is rather impressive. At one time the group may be preparing a catalog of military antennas, compiling a comprehensive bibliography of peanut technology, and collecting information on the shape of the earth. At another time it may be preparing a manual of small-scale cannery operation, a study of dipilatories, a review of the literature on thin metal films, and an examination of the factors which influence the location of various manufacturing enterprises.

This diversity of interest is typical of TIS activities. The personnel qualify as authorities on information retrieval, but they also have worked on such unlikely subjects as two-phase fluid-flow, charcoal, patents, bamboo, insect repellants, engineering education, dipilatories, and wood preservatives. From the standpoint of education, the staff is strongest in chemical engineering. But, variety is the keynote of TIS.

MOST of the work which TIS performs is some form of a literature survey or a related service. The related services include advice or assistance in setting up an information file and aid in finding information. At times the greatest service, especially for industry or other off-campus organizations, may be simply obtaining copies of publications. It is possible to obtain borrowed or photo-duplicated reports, books, etc., from Georgia Tech even though they are available only in other libraries or even in other countries.

Another sideline service which is often very useful is aid in locating translators and translations. Local translations are usually performed by students, graduate and undergraduate, from foreign countries. Usually they are advanced students who consider the language being translated as their native tongue.

MOST literature surveys can be classified in one of the following types:

1. *Answers to specific questions*
2. *Abbreviated bibliographies*
3. *Bibliographies*
4. *Annotated bibliographies*
5. *Literature reviews*
6. *Continuous surveys*
7. *Patent surveys*

Type 1, the simplest, may be short answers to one or a few short technical questions. Often such questions can be answered very quickly, but their value to the inquirer will be quite great. Sometimes these simple questions seem ridiculous: "What is the frequency of the sound impulses emitted by a shrimp?" or "How can we keep our candles from sticking in the mold?". Neither was funny, however, to the groups interested in the answers to these questions.

Type 2 is abbreviated bibliographies. They are called for when it is necessary

Photo Credit: Bill Diehl Jr.

GEORGIA TECH'S TECHNICAL INFORMATION SECTION SEARCHES REPORTS.

to compile a list of references in order to substantiate the findings. Pertinent references are particularly necessary when individual authorities disagree on some points.

Type 3, bibliographies, are generally considered rather comprehensive. Occasionally, they cover a specified number of years, or they may contain only the information published in certain journals or reviewed in a given abstract journal. In all cases, they represent a relatively complete collection of the literature on the given subject. Often they include hundreds of individual references, including books, journal articles, trade literature and catalogs, and patents.

Type 4 is annotated bibliographies. It is sometimes desirable to have an extensive collection of all the literature which can be found, together with an indication of the exact material contained in each reference. The Engineering Experiment Station has published book-length annotated bibliographies on such subjects as the shape of the earth, two-phase gas-liquid flow, peanut technology, frozen foods, manganese dioxide for batteries, and solvent extraction. Annotated bibliographies are not necessarily long, for it is often desirable to have abstracts with even the most superficial bibliographies.

There are two basic types of abstracts, indicative and informative. Indicative abstracts simply indicate the type of information to be found. Informative abstracts, however, serve as condensations of the original information. Informative abstracts average about one tenth as long as the original and cost from $2.50 to $18.00 each, depending on the length and nature of the original. The usual cost is about $7.00. Informative abstracts are generally smaller and much less expensive.

It is often possible to obtain permission to copy the brief summaries or abstracts which many journals print with their articles. Some abstracting journals, like *Chemical Abstracts,* sell the right to reproduce their abstracts. Unfortunately, the price is quite steep, although it is very

inexpensive compared to the cost of preparing similar abstracts yourself.

Type 5, one of the most useful compilations of references, is literature reviews. In preparing a review, it is customary to collect all the literature which can be made readily available, although the survey can be as extensive or as superficial as fits the individual need. Usually, the original references are read and digested, and a summary is drawn of the findings of the various investigators who have written on the subject. TIS recently published a review on the subject "Solubility of Solids in Liquids at High Pressures." Another extensive review prepared by TIS was the review accompanying the annotated bibliography of "Two-phase Gas-liquid Flow."

Type 6, continuous literature surveys, is aptly described by its designation. Many companies have their librarians or other persons on their staff perform continuous literature surveys. It is not uncommon for a company or a group of companies to have a continuous review prepared for them by an external organization. At present, TIS conducts a continuous survey, on a monthly basis, in which references from more than 300 journals are included. Most of the branches of the Engineering Experiment Station make use of this monthly literature review. Analog computers is the subject of another survey prepared for campus personnel. Fifty or more analog computer groups throughout this country and abroad have learned of the survey and requested to be added to the mailing list. The Station has found these abstracts so valuable for its own research staff that it is now considering offering *tailor made* continuous literature surveys to interested companies.

Type 7, patent surveys, vary to an extraordinary degree in their complexity. At the beginning of each patent application, it is customary to have a patent search made. These searches are usually performed by patent attorneys for a nominal fee. They include only the most obvious references related to the patent idea. However, they often serve to show that the idea has already been patented and that the investigator would be wise not to apply. In other cases, patent surveys are made in an effort to determine how a given product is being manufactured. If a process is known to be patented, it is conceivable that a patent survey will lead to the patent which gives a complete description of the process. Obviously, the patentee is still protected by his patent. However, the knowledge of a patented process may lead a competitor to a research investigation which will produce an even better process.

The most extensive literature surveys are patent searches called *validity searches*. When a company is accused of infringing on a patent, it may undertake an extremely comprehensive review in the hope that it will be able to discover some early reference to the patented invention. If a reference precedes the date of the patent, it may indicate that the patented idea is not a patentable invention but an adaption of earlier thinking. Thus, the competitor can break the patentee's monopoly and the patented subject will become available to the general public. Unfortunately, this procedure may sometimes be used to defeat deserving patents as well as unjust ones.

The relentless advance of technology will fill our libraries with new information, at an ever increasing rate, to the ultimate benefit of the entire world. It is obvious that the maximum benefit can only be achieved if this new information is used wisely.

The South is fortunate, and Georgia especially so, to have the basic ingredients necessary for continued industrial expansion: favorable geographic location, rapidly expanding markets, extensive manpower resources, and experimental research to fill in the gaps of scientific knowledge. Such reservoirs as the excellent Georgia Tech Library can best be tapped by a group of trained, experienced information scientists. The Technical Information Section of the Engineering Experiment Station provides this capability for Georgia Tech and the South.

Edited in Retrospect

At some conferences many enriching statements can be heard, while at others mediocrity of thought exists. The following quotes represent an enriching conference.

Why send a man into outer space, when machines can obtain the same information? Charles Lundquist, space authority of the George C. Marshall Space Flight Center says, "There will be so many satellites in outer space in ten years, that we may need a maintenance man to make the rounds."

"A volume of ova equal to a gallon plus a volume of sperm the size of an aspirin produced 2.7 billion people," states nationally recognized biochemist, Dr. Philip Handler.

"Investigation of atoms is analogous to bouncing baseballs off an object and seeing how the baseball bounces off," explained Dr. Hugh Wolf, Director of Publications, American Institute of Physics.

Mr. William M. Pinkerton, News Officer for Harvard University, says, "The good public information man knows the outside better than the insiders and he knows the inside better than the outsiders. But, he doesn't know the outside as well as the outsiders nor the inside as well as the insiders."

Dr. Warren Weaver, now Vice-president of Alfred P. Sloan Foundation, proclaims: "The public must know and understand science not only because it is interesting, not because it's wonderous, not because it's a part of education, but because its understanding is essential for correct decisions to be made in our technological society.

These statements were made by the outstanding scientists and public information officers attending The Southern Regional Science Seminar. Four enlightening and informative days were the result of this activity designed to broaden and develop the background and understanding of science for college and university public information officers. The conference held in February also aided the participants by increasing their skills in communicating science news to the public. Dr. James E. Boyd and Bryan W. Miller were Georgia Tech's representatives at the Conference.

g

n
.
.
s
,
t
r
n

e
-
p

e-
es
s,

d
n
vs
w
ns

P.
nd
ot
u-
r-

ci-
he
nd
ed
ng
on
he
ng
an
he